

# Time-Dependent Density-Functional Approach for Biological Chromophores: The Case of the Green Fluorescent Protein

Miguel A. L. Marques,<sup>1</sup> Xabier López,<sup>2</sup> Daniele Varsano,<sup>1,3</sup> Alberto Castro,<sup>4,3</sup> and Angel Rubio<sup>1,3,5</sup>

<sup>1</sup>*Departamento de Física de Materiales, Facultad de Químicas, UPV/EHU, 20018 San Sebastián, Spain*

<sup>2</sup>*Departamento de Química, Facultad de Químicas, UPV/EHU, 20018 San Sebastián, Spain*

<sup>3</sup>*Donostia International Physics Center (DIPC), 20018 San Sebastián, Spain*

<sup>4</sup>*Departamento de Física Teórica, Universidad de Valladolid, Valladolid, Spain*

<sup>5</sup>*Centro Mixto CSIC-UPV/EHU, 20018 San Sebastián, Spain*

(Received 8 January 2003; published 23 June 2003)

We performed first-principles calculations of the optical response of the green fluorescent protein (GFP) within a combined quantum-mechanical molecular-mechanics and time-dependent density-functional theory approach. The computed spectra are in excellent agreement with experiments assuming the presence of two, protonated and deprotonated, forms of the photoreceptor in a  $\sim 4:1$  ratio, which supports the conformation model of photodynamics in GFP. Furthermore, we discuss charge transfer, isomerization, and environment effects. The present approach allows for systematic studies of excited-state electron-ion dynamics in biological systems.

DOI: 10.1103/PhysRevLett.90.258101

PACS numbers: 87.15.Mi, 71.15.Qe, 78.40.-q, 87.15.Aa

Not surprisingly, the theoretical understanding of biophysical processes is a very active field of research. In particular, there have been spectacular advances in the characterization of structural and dynamical properties of complex biomolecules by a combination of quantum-mechanical and classical-molecular mechanics methods [1,2]. However, and in spite of the large amount of experimental work in photoactive molecules, the theoretical description of the interaction of these molecules with external time-dependent fields is very much in its infancy. Many biological processes rely on a subtle interplay between optical absorption in the photoactive center and its coupling to internal vibrational modes and to the environment (hosting protein and solvent)—e.g., vision, that is triggered by a photoisomerization mechanism. Another paradigmatic case is the green fluorescent protein (GFP). This molecule has become a unique tool in molecular biology due to its fluorescence and inertness when attached to other proteins [3]. The chromophore consists of chemically modified Ser65, Tyr66, and Gly67, which form two consecutive rings, the phenol-type ring of Tyr66 and a five-membered heterocycle (imidazolinone) formed by the backbone of Tyr66, the carbonyl carbon of Ser65, and the nitrogen of the backbone of Gly67 (see Fig. 1). The tyroxyl hydroxyl group is part of a complex hydrogen bond network that depending on the environment favors its protonated (neutral) or deprotonated (anionic) form.

The optical absorption spectrum of the wild type (wt)-GFP, measured at 1.6 K, shows two main resonances at 2.63 and 3.05 eV [4] that are attributed to the two thermodynamically stable protonation states of the chromophore (negative and neutral configurations, respectively). The equilibrium between those two states can be controlled by external factors such as pH and mutations that affect the chromophore environment [4–6]. Excitation at

either frequency leads to fluorescent green-light emission, peaked at 2.44 eV, which is the main mechanism for energy release in wt-GFP. This internal photoconversion process occurs very rapidly by excited-state proton transfer [7] for the neutral chromophore. Only recently the optical absorption spectrum of the GFP anion chromophore was measured *in vacuo* [8]. It consists of a main peak at 2.59 eV in very close agreement with the peak assigned to the anionic configuration of the wt-GFP. It is clear that the photophysics of the GFP is governed by a complex equilibrium between the neutral and anionic configurations. The relevance of other proposed configurations has been ruled out [9,10]. Several calculations of

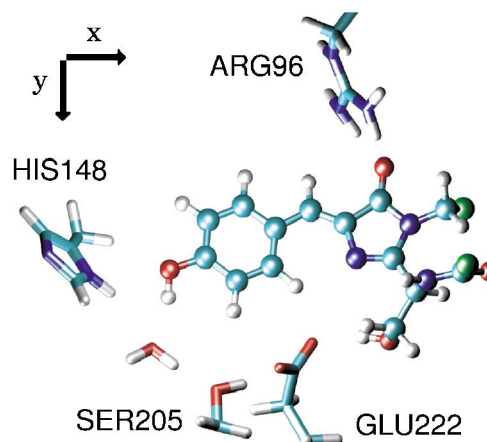


FIG. 1 (color). Optimized structure of the neutral chromophore and its closest charged residues inside the IGFL protein: His148, Arg96 (positive), and Glu222 (negative). Green spheres correspond to the H-link atoms for the QM/MM procedure (see text). Note that the dipole created by Arg96 and Glu222 does not have a component in the relevant  $x$  direction.

the excited states have been performed using *ab initio* quantum-chemistry methods [9–11]. Unfortunately, these calculations yielded optical absorption spectra in not satisfactory agreement with experiment. A microscopic model accounting for the optical properties of the GFP chromophore including its blinking and photobleaching dynamics is still missing [7]. In the present work we prove that a combination of a quantum-mechanics/molecular mechanics method (QM/MM) to obtain the structural properties and time-dependent density-functional theory (TDDFT) [12] to treat the electronic excitations provides a very good description of the optical properties of the GFP *in vacuo*.

Over the last years TDDFT has emerged as a very successful framework to describe the optical spectra of nanostructured materials [12]. It has been shown to work very well for the description of electronic excitations of first and second row elements as well as noble and early transition metal systems. Therefore, we think it is a promising tool for the description of excited-state biological processes. In practice, we used TDDFT to obtain optical absorption spectra as implemented by some of us in OCTOPUS [13]. All the dynamical quantities are computed by evolving the electronic wave functions in real time and real space, as first introduced in Ref. [14]. This approach does not rely on perturbation theory and is competitive with implementations in the frequency domain [15]. The electron-ion interaction is described through norm-conserving pseudopotentials [16], and exchange-correlation effects were treated within the local-density approximation (LDA). In the present study the specific form of the exchange-correlation functional should not introduce major modifications [17]. This is, however, untrue for studies of charge-transfer processes and isomerization mechanisms, where the (incorrect) asymptotic behavior of the LDA potential might play a role [12]. We used a uniform grid spacing of 0.23 Å and a time step of 0.0017 fs, which ensure the stability of the time-dependent propagation and yield spectra with a resolution better than 0.1 eV.

The GFP structures were prepared according to the following protocol. The x-ray structure (1.9 Å resolution) of the GFP from Ref. [18] (PDB code: 1GFL) was taken as the reference structure. The protein is composed of 238 amino acids, the last 9 C terminal residues being missing from the x-ray structure. It is folded in a  $\beta$ -sheet barrel conformation with the chromophore in its interior. In the preparation of the structures, we kept only the crystallographic water molecules that lied within 6 Å of the chromophore. The position of the hydrogens was initially obtained using the HBUILD command [1] of CHARMM [19]. In order to relax the geometry of the system three consecutive optimizations were carried out. An initial optimization was performed with all backbone and chromophore atoms fixed at their crystallographic positions. Nonbonded interactions were treated using a spherical cutoff of 12 Å with an atom-based

potential shift method for the van der Waals terms and the force-shift method for the electrostatic terms. In a second step we allowed only relaxation of the coordinates of the chromophore. As the GFP chromophore (especially the heterocycle) does not correspond to the typical structure adopted by serine, tyrosine, or glycine, there is a lack of specific force field parameters. Therefore, we performed this step within a QM/MM hybrid method [20], with an AM1 semiempirical Hamiltonian [21] to describe the quantum subsystem. The QM region was formed by three amino-acid sequences, Ser65, Tyr66, and Gly67, and the frontier between QM and MM regions was treated within the H-link approximation. In this approach a H atom is included whenever the frontier between the QM and MM regions passes through a chemical bond. This H-link atom is forced during the minimization to be aligned with the frontier bond and does not interact with the MM atoms. In order to avoid the QM/MM frontier to be in the C(O)-N peptide bond, the carbonyl group of Gly67 was removed from the QM subsystem and the carbonyl group of Phe64 was included. In this way, the 2 H-link atoms cut through C-C bonds. Finally, from this geometry we performed a full relaxation of the chromophore and every side-chain atom within 10 Å. The final structure of the chromophore with the most important neighbor residues is depicted in Fig. 1.

The anionic form of the chromophore was prepared from deprotonation of the Tyr66 and protonation of Glu222. The proton transfer is mediated through a water molecule and Ser205. The water molecule, depicted in Fig. 1, receives a proton from Tyr66 and donates a proton to Ser205. Ser205 in turn donates it to Glu222, which is therefore neutralized. The role of the protein backbone is important for the structural relaxation of both the neutral and the anionic form. The main structural difference to the neutral form is the relative orientation of the Tyr66 ring with respect to the five-membered heterocycle: In the neutral form the two ring planes are found to be slightly displaced from planarity, the  $C_\alpha$ - $C_\beta$ - $C_\gamma$ - $C_\delta$  dihedral being  $-10.9^\circ$  (in the gas phase the planar configuration is the stable one). In the anionic form the coplanarity of the two rings is enhanced, through the reduction of the  $C_\alpha$ - $C_\beta$ - $C_\gamma$ - $C_\delta$  dihedral to  $-0.9^\circ$ . Concerning bond distances, the largest change upon deprotonation of Tyr66 corresponds to the shrinking of the C-O bond distance from 1.39 Å in the neutral form to 1.30 Å in the anionic form.

From these relaxed structures, we extracted the quantum-mechanical subsystem that we subsequently used for our TDDFT calculations. The computed spectra of the neutral and anionic conformations of the GFP chromophore are given in Fig. 2, together with the experimental spectrum of the wt-GFP. The spectra are averaged over the three spatial directions. Light polarized along the x direction is responsible for the lowest optical transition (see the inset of Fig. 2) and corresponds to a

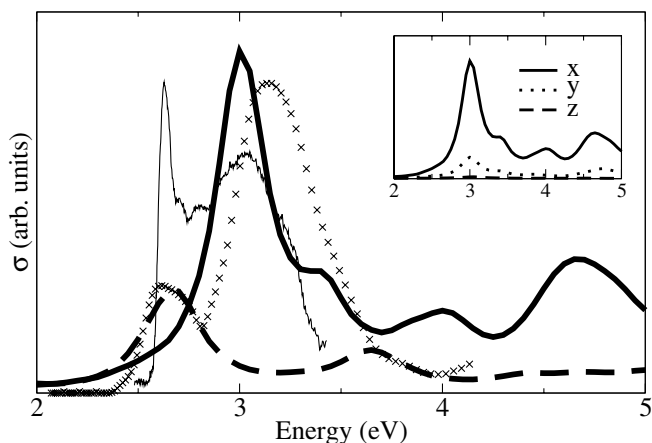


FIG. 2. Computed photoabsorption cross section of the neutral (thick solid line) and anionic (thick dashed line) chromophores, along with experimental results taken from Refs. [4,8] (thin solid line and crosses, respectively). For comparative purposes, we divided the anionic results by 4 with respect to the neutral results. Inset: decomposition of the computed spectra of the neutral chromophore in the three directions, showing the inherent anisotropy of the GFP molecule.

$\pi$ - $\pi^*$  transition between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) of both neutral and anionic forms. The molecule is nearly transparent to visible light polarized along the other two orthogonal directions. GFP turns out to be a rather anisotropic molecule in the visible, property that can be used to enhance the photodynamical processes in well-oriented GFP samples for optoelectronic devices. The differences between the LUMO and HOMO eigenvalues, sometimes taken as an estimate for the HOMO-LUMO excitation energy, are, respectively, 2.19 and 1.61 eV, for the neutral and anionic structures. It is common practice to use these values as estimations of the physical excitations. This leads, in the case of the GFP, to a very bad agreement with experiment. However, in TDDFT the difference of one-particle eigenvalues is renormalized by the Coulomb and exchange-correlation terms [12]. Once these effects are included, the main excitation peaks for the neutral and anionic forms move to 3.01 and 2.67 eV. These values are now in really good agreement with the measured excitation energies, located at 3.05 and 2.63 eV [4]. The  $\pi$ - $\pi^*$  excitations are no longer pure HOMO-LUMO transitions and include contributions from virtual particle-hole excitations involving several occupied and unoccupied states. Some of the relevant states are depicted in Fig. 3. The oscillator strength of the  $\pi$ - $\pi^*$  transition is larger in the anionic than in the neutral GFP. It is, however, possible to obtain a quantitative description of the spectra of the wt-GFP by assuming a  $\sim 4:1$  ratio for the concentration of the neutral/anionic forms. This ratio is very close to the estimated experimental ratio of 80% neutral and 20% anionic [3]. The measured peaks can be clearly assigned to either the neutral or the anionic forms of the GFP.

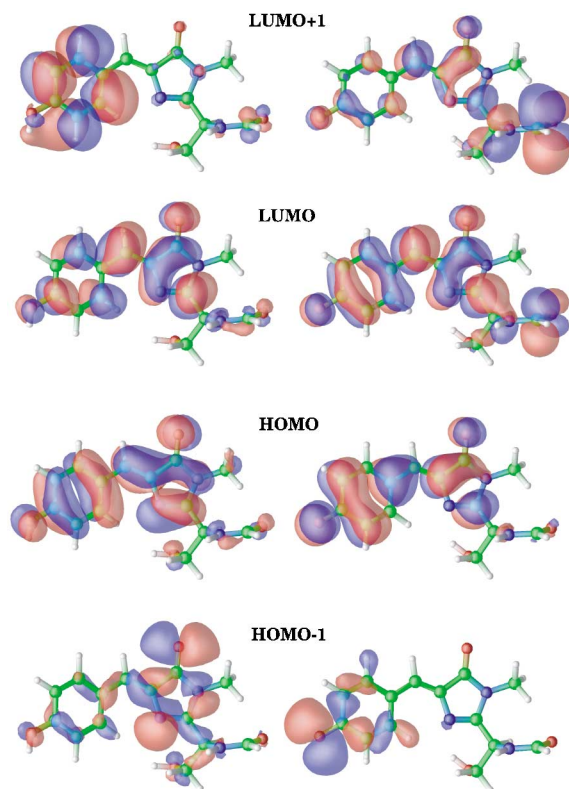


FIG. 3 (color). Kohn-Sham wave functions of the neutral (left) and anionic (right) GFP chromophores (red +, blue -). These are the most important states involved in the main collective excitation of Fig. 2, although the contribution of several other occupied and unoccupied states is not negligible.

The suggested three-state model of the wt-GFP photo-physics contains the two thermodynamically stable neutral and anionic configurations, and an unstable intermediate state that apparently corresponds to the anionic chromophore in a nonequilibrium protein environment [5]. Other intermediate state relaxation processes through cis-trans isomerization and a new protonated zwitterionic form have also been proposed [9]. Our calculations give further support to the predominance of the neutral and anionic forms in wt-GFP in agreement with the analysis of the infrared spectra [10]. This sustains the proposed proton-shuttle mechanism between the protonated and deprotonated forms, through the corresponding charged states of the Glu222 residue where the proton-shuttle ends. Furthermore, the structural differences between the two anionic structures is very small (changes the main absorption peak by  $\sim 0.1$  eV). We have also computed the cationic conformation derived from the neutral form by protonation on the heterocyclic free-radical N. Its spectrum consists of a major peak at 2.6 eV along with a shoulder at 3.05 eV in relative agreement with the measured value for a slightly different model chromophore in aqueous solution, 3.15 eV [22]. This predictive power can prompt the use of TDDFT combined with time-resolved optical spectroscopy to map the concentration of the different GFP forms in solution upon

excitation and extract information on the isomerization paths and lifetimes of the initial and transition states.

The fact that the computed absorption spectra of the isolated chromophore is in close agreement with the measured spectra points to an efficient electrostatic shielding of the chromophore by the rigid  $\beta$  barrel. This justifies the comparison between *in vacuo* and *in vivo* spectra [8]. In spite of the rigidity of the  $\beta$  barrel, there is a change in the neutral/anionic concentration ratio with temperature, glycerol content of the solution, and external pH [3]. This fact can be understood by considering that the Tyr66 hydroxyl points toward the protein surface and it is H bonded to a water molecule, and therefore sensitive to bulk proton concentrations. In particular the raise of the anionic population as the pH is increased can be explained by the protonation Glu222 [23]. This neutralization triggers the conversion from neutral to anionic isomers in solution. In the same way, our calculations show that the excitation leads to a slight charge transfer from the phenolic to the imidazolinone ring (mainly at the bridging group as already observed in previous calculations [11]). We observed that the main effect of the excitation in the structure is to increase the strength of the single bonds and weaken the double bonds. This has important consequences for the proton affinity of the different sites. Clearly, after the  $\pi$ - $\pi^*$  transition in both neutral and anionic forms, the density along the heterocyclic N increases, thus enhancing its proton affinity (see states in Fig. 3). There is also a reduction of the charge in the phenolic O atom of the Tyr66 after excitation of the neutral GFP that may trigger the deprotonation.

In conclusion, we have shown that a combined QM/MM and TDDFT approach is able to reproduce the optical response of the GFP. This is a major step toward the first-principles description of the excited-state dynamic of important biological photoreceptors. Transient and time-resolved optical spectroscopy could be studied. However, in spite of the good agreement, some questions remain open. For example, how does the excitation in the GFP trigger the proton-shuttle mechanism and what is the time scale for this process? To answer these questions we have to go beyond the present work where only the QM subsystem was considered to obtain the optical response. The proton transfer involves structural modifications of the environment that have to be properly described by a modified QM/MM approach. The standard QM/MM methods [2] have to be generalized to cope with excited-state dynamics, and eventually to include the possibility of charge transfer between the QM to the MM parts. Work along these lines is in progress.

We thank G. F. Bertsch, F. Cossio, J. Ugalde, and P. M. Echenique for enlightening discussions and support. We thank the computing facilities of CEPBA. This work was supported by the UPV, MCYT (MAT 2001-0946),

and the European Community RTN (HPRN-CT-2000-00167).

- 
- [1] A. T. Brunger and M. Karplus, *Proteins* **4**, 148 (1988).
  - [2] A. Laio, J. VandeVondele, and U. Rothlisberger, *J. Chem. Phys.* **116**, 6941 (2002), and references therein.
  - [3] K. F. Sullivan and S. A. Kay, *Green Fluorescent Proteins* (Academic Press, San Diego, 1999).
  - [4] T. M. H. Creemers *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2974 (2000); *Nat. Struct. Biol.* **6**, 557 (1999).
  - [5] K. Brejc *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2306 (1997).
  - [6] U. Haupts *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13 573 (1998).
  - [7] P. Schwille *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 151 (2001); R. A. Cinelli *et al.*, *Phys. Rev. Lett.* **86**, 3439 (2001).
  - [8] S. B. Nielsen *et al.*, *Phys. Rev. Lett.* **87**, 228102 (2001).
  - [9] W. Weber *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6177 (1999).
  - [10] H. Y. Yoo *et al.*, *J. Phys. Chem. B* **105**, 2850 (2001); V. Tozzini and R. Nifosi, *J. Phys. Chem. B* **105**, 5797 (2001).
  - [11] V. Helms, C. Winstead, and P. W. Langhoff, *J. Mol. Struct.* **506**, 179 (2000); A. A. Voityuk, M. E. Michel-Beyerle, and N. Rösch, *Chem. Phys.* **231**, 13 (1998).
  - [12] M. A. L. Marques and E. K. U. Gross, in *A Primer in Density Functional Theory*, edited by C. Fiolhais, F. Nogueira, and M. A. L. Marques (Springer, Berlin, 2003); G. Onida, L. Reining, and A. Rubio, *Rev. Mod. Phys.* **74**, 601 (2002), and references therein.
  - [13] M. A. L. Marques *et al.*, *Comput. Phys. Commun.* **151**, 60 (2003) and [www.tddft.org/programs/octopus](http://www.tddft.org/programs/octopus).
  - [14] K. Yabana and G. F. Bertsch, *Phys. Rev. B* **54**, 4484 (1996); *Int. J. Quantum Chem.* **75**, 55 (1999).
  - [15] A. Rubio *et al.*, *Phys. Rev. Lett.* **77**, 247 (1996); *Int. J. Mod. Phys. B* **11**, 2727 (1997).
  - [16] N. Troullier and J. L. Martins, *Phys. Rev. B* **43**, 1993 (1991); L. Kleinman and D. M. Bylander, *Phys. Rev. Lett.* **48**, 1425 (1982). In this work we used the core radii  $r_c = 1.49, 1.39$ , and  $1.39$  a.u. for the  $s, p$ , and  $d$  components of the carbon, nitrogen, and oxygen pseudopotentials.
  - [17] M. A. L. Marques, A. Castro, and A. Rubio, *J. Chem. Phys.* **115**, 3006 (2001).
  - [18] F. Yang, L. G. Moss, and G. N. Phillips, Jr., *Nature Biotech.* **14**, 1246 (1996).
  - [19] B. R. Brooks *et al.*, *J. Comput. Chem.* **2**, 187 (1983); A. D. MacKerell, Jr. *et al.*, *J. Phys. Chem. B* **102**, 3586 (1998).
  - [20] M. J. Field, P. A. Bash, and M. Karplus, *J. Comput. Chem.* **11**, 700 (1990).
  - [21] M. J. S. Dewar *et al.*, *J. Am. Chem. Soc.* **107**, 3902 (1985).
  - [22] A. F. Bell *et al.*, *Biochemistry* **39**, 4423 (2000).
  - [23] Replacing the Ser65 by several mutants can change the H network, preventing ionization of the Glu222 and suppressing the anionic peak [5]. Similarly, a reduction of the pH leads to a drop of the neutral peak [6].